Active Learning and Collaborative Filtering

Valdemaras Repsys  
Francesco Ricci  
Mehdi Elahi  

Free University of Bozen-Bolzano, Italy

1. Introduction

Choosing the right product to consume or purchase is nowadays a challenging problem due to the growing variety of eCommerce services and the informational globalization. Recommender Systems (RSs) aim at addressing this problem providing personalized suggestions for digital content, products or services. In this paper we are concerned with collaborative filtering (CF) RSs (Koren, 2008); they use item ratings provided by a population of users to predict unknown ratings. In fact, the rating prediction accuracy and consequently the quality of the recommendations (i.e., items with the largest predicted ratings) depends not only on the characteristics of the prediction algorithm, but also on the available ratings. The more ratings are available and the more information they bring, the higher the recommendation accuracy is. Therefore, it is very important for a RS to keep acquiring new and useful ratings in order to maintain and even improve the quality of the recommendations.

2. Elicitation Strategies

Different techniques, defined as rating elicitation strategies, can be exploited for choosing the items to be presented to the user for rating. We defined and tested some “pure” strategies, i.e., implementing a straightforward heuristic, e.g., asking to the (simulated) user to rate the selected recommendations (if he has experienced them), but also strategies that we called “partially randomized”, which, in addition to asking to the user to rate the items selected by a “pure” strategy, they require him to rate some randomly selected items as well. We believe that partially randomized strategies better represent the real-world elicitation process, because in real RSs ratings are added not only by actively asking the user to rate items, but users might also rate some (random from the system perspective) items while browsing or searching the catalog on their own initiative.

3. Methodology

We have evaluated a set of ratings elicitation strategies. Some of them have been proposed in a previous work (Rashid et al., 2002) (popularity, random and variance), and some new ones that we defined as prediction-based strategies: binary-prediction, highest-predicted, lowest-predicted, highest-lowest-predicted. Highest-lowest-predicted, for instance, asks the
user to rate items with the lowest and highest predicted ratings. In order to study the effect of these elicitation strategies we set up a simulation procedure. The goal was to simulate the evolution of a RS’s performance exploiting these strategies. In order to run such simulations we partition all the available (not null) rating data into three different matrices with the same number of rows and columns: $K$(known), $X$(unknown) and $T$(test). $K$ contains the ratings known by the system, $X$ contains the ratings known by the users that are elicited by the strategies and $T$ contains ratings used for testing the system performance. The MovieLens and Netflix rating databases were used for our experiments. From the full Netflix data set we extracted the first 100,000 ratings entered into the system. So, we built our own offline experimental environment creating a software which simulates the real process of rating elicitation, rating database growth, and system adaptation (retraining) to the new set of data. We considered four evaluation measures: mean absolute error (MAE), precision, coverage and normalized discounted cumulative gain (NDCG). For computing precision we extracted, for each user, the top 10 recommended items (whose rating values appear in $T$) and considered as relevant the items with rating values (in $T$) equal to 4 or 5. The coverage of a recommender system is measured as the proportion of the full set of items over which the system can form predictions or make recommendations. Normalized discounted cumulative gain (NDCG) measures how close the ranking of the items predicted by the RS is close to the optimal ranking.

4. Results

The evaluation showed that different strategies can improve different aspects of the recommendation quality and in different stages of the rating database development. Moreover, the pure strategies may incur in the risk of increasing MAE if they keep adding only ratings of a certain type, e.g. those with the highest values, as it is the case for highest-predicted strategy that is an approach often adopted by real RSs. In conclusion from these experiments, one can conclude that the selection strategy must be matched to the evaluation measure. In fact, there is no single best strategy, among those that we evaluated, that dominates the others for all the evaluation measures. The random strategy is the best for NDCG, whereas for MAE and precision we would suggest using lo-high predicted, performing quite well for both measures. In addition, prediction-based strategies neither address the problem of new users, nor of new items. Popularity and variance strategies are able to select items for new users, but can not select items that have no ratings.

References
